Coordination Chemistry Reviews
Elsevier Publishing Company Amsterdam – Printed in The Netherlands

TWO PROBLEMS INVOLVED IN SOLVING COMPLEX FORMATION EQUILIBRIA: THE SELECTION OF SPECIES AND THE CALCULATION OF STABILITY CONSTANTS

A. VACCA, A. SABATINI and M. A. GRISTINA

Laboratorio CNR Istituto di Chimica Generale dell'Università di Firenze
Via J. Nardi 39, 50132 Florence (Italy)

For the investigation of systems containing several complex species at equilibrium, a knowledge of the stoichiometry of the complexes and of their degree of formation is important. It is necessary to determine the type of species existing and to calculate their stability constants. These two problems are closely interconnected. Only in rare examples of simple equilibria is it possible to characterize the existence of different complexes, using experimental data which are not dependent on the stability constants. On the other hand, a chemical model (number and type of species which can be formed in the equilibrium mixture) is always needed for the calculation of the stability constants.

The method commonly employed is to assume arbitrarily that the chemical model is known. This assumption is generally supported by the fact that the species taken into consideration have sometimes been isolated as solid phases. Alternatively, analogy with apparently similar systems may be made. On the basis of this hypothesis, it is possible to calculate the formation constants using one of the computer methods, which are by now widely applied. Apart from some complications which may arise during the calculation (non-convergence, negative values for some constants, etc.), one obtains, after some cycles, the stability constants of the assumed species and their standard deviations. At this point a statement is generally made about the agreement between experimental and calculated data. The fit may be poor, satisfactory or good. These statements, however, are often unfounded, since a sufficiently rigorous statistical basis is missing. Should the agreement be considered unsatisfactory, new species are introduced in the calculation, in order to obtain a better fit with the experimental data. If a better agreement is found, the new species are retained in subsequent calculation. However, in this case also, sound statistical arguments should show that the new hypothesis is significantly better than the preceding one.

The present communication deals with a new procedure for investigating complex formation equilibria and with the application of some statistical tests which are useful in the choice of the most probable chemical model for the system under consideration.

A CALCULATION OF THE STABILITY CONSTANTS

The application of electronic computers to the determination of stability constants has been recently reviewed [1] The most commonly used procedures employ one of two mathematical approaches, i e the Gauss–Newton least squares method[2] or the pit-mapping method[3] In both cases, an observable variable quantity y, which is a function of the equilibrium constants β_i

$$y = f(\beta_i) \tag{1}$$

is chosen, and the sum of the squares of the weighted residuals

$$U = \sum_k (y_k^{\text{obs}} - y_k^{\text{calc}})^2 w_k \tag{2}$$

is minimized with respect to β_i, as adjustable parameters In eqn (2), y_k^{obs} represents the result of the kth measurement of the quantity y, y_k^{calc} the corresponding value calculated according to eqn (1) and w_k the weight associated with the kth measurement

We have recently developed a new method for least squares refinement of stability constants, which has been applied successfully to one metal–one ligand systems in aqueous solution, using pH-titration data This method allows the simultaneous determination of the formation constants of all complexes present in solution, such as simple, hydroxy and protonated species, both mono- and polynuclear

Each point on a titration curve represents an equilibrium state among N different species of general formula $H_pM_qL_r$ (for details on the symbolism, see ref 4) The concentration of each species $H_pM_qL_r$ in the kth point is given by

$$[H_pM_qL_r]_k = \beta_{pqr} [H]_k^p [M]_k^q [L]_k^r$$

where β_{pqr} is the formation constant of the species under consideration, and $[H]_k$, $[M]_k$ and $[L]_k$ are the concentrations of the hydrogen ion, of the free metal and of the free ligand, respectively

The mass balance equations for the kth point are

$$T_{k,1} = [H]_k + \sum_{i=1}^{N} p\,\beta_{pqr} [H]_k^p [M]_k^q [L]_k^r \tag{3}$$

$$T_{k,2} = [M]_k + \sum_{i=1}^{N} q\,\beta_{pqr} [H]_k^p [M]_k^q [L]_k^r \tag{4}$$

$$T_{k,3} = [L]_k + \sum_{i=1}^{N} r\,\beta_{pqr} [H]_k^p [M]_k^q [L]_k^r \tag{5}$$

where $T_{k,1}$, $T_{k,2}$ and $T_{k,3}$ are the analytical concentrations of hydrogen ion, metal ion and ligand, respectively These equations are non-linear in the unknowns $[M]_k$, $[L]_k$ and β_{pqr}

The value of $[H]_k$ is obtained from the potentiometric measurement.

The number of equations, in a set of n points, is $3n$, and the number of unknowns is $2n + m$, where m is the number of formation constants to be determined. The degree of freedom of the system is then $n - m$.

In the present method, the error square sum

$$U = \sum_{k=1}^{N} \sum_{j=1}^{3} (T_{k,j}^{\text{calc}} - T_{k,j}^{\text{obs}})^2$$

is minimized with respect to the parameters $[M]_k$, $[L]_k$ and β_{pqr} by the Newton–Raphson least squares method [5]. The residuals of the mass balance equations, estimated by considering random errors in the analytical concentrations of the reactants, are all of the same order and are far greater than those from experimental potentiometric data. For this reason, the use of equal weights is satisfactory. The corrections of the unknown parameters, obtained from a least squares cycle, are multiplied by a factor such that the decrease of U is maximum for that cycle [6]. The iterative procedure is continued, using the new approximation of the parameters, until all the shifts are less than a pre-established fraction of the values of the parameters to be refined.

The present method differs from those previously described [2,3] because

(i) the concentrations of free metal ion and free ligand for each point are considered as independent variables, at the same level as the stability constants. In the other refinements only the stability constants are treated as adjustable parameters, while $[M]_k$ and $[L]_k$ are calculated by simultaneous solution of mass balance equations (4) and (5).

(ii) the Newton – Raphson least squares method is employed, and it has been found more efficient than the conventional Gauss – Newton method. When a situation of non-convergence occurs (for example, due to poor estimation of the starting values of β_{pqr}), the refinement procedure continues, for a pre-established number of cycles, using the *steepest descent* method [7], and

(iii) the calculated shifts of the parameters are optimized: a similar procedure has recently been applied successfully to the calculation of vibrational force constants [8].

We have obtained satisfactory results for all the systems investigated so far with a high rate of convergence, even if the starting values were estimated roughly. Furthermore, using this method, the calculated residuals follow a nearly normal distribution with a mean value close to zero. This allows us to make statistical decisions about the goodness of the fit and to check whether the chemical model adopted for the calculation of constants is a good representation of the data.

B. TESTS OF SIGNIFICANCE

As a measure of the agreement between experimental and calculated data and by the analogy with crystallographic calculations, we can take the well known R factor defined as

$$R = \sqrt{\frac{\sum_{i=1}^{3n} (f_i^{\text{calc}} - f_i^{\text{obs}})^2 w_i}{\sum_{i=1}^{3n} (f_i^{\text{obs}})^2 w_i}}$$

where f_i^{obs} are the experimental quantities $T_{k,1}$, $T_{k,2}$ and $T_{k,3}$ (defined in the preceding section), f_i^{calc} the corresponding values calculated according to eqns (3)–(5), and w_i the weights assigned to each mass balance equation. We can define another quantity

$$R_{\text{lim}} = \sqrt{\frac{\sum_{i=1}^{3n} \epsilon_i^2 w_i}{\sum_{i=1}^{3n} (f_i^{\text{obs}})^2 w_i}}$$

where ϵ_i is the residual in the ith equation calculated from pessimistic estimates of the errors in all the experimental quantities and using the usual rules for propagation of error. This quantity can be regarded as a significance limit for the function R and, as a consequence, if

$$R < R_{\text{lim}}$$

we shall be willing to accept the hypothesis that the fit of the data is satisfactory. If, on the contrary, the computed value of the R factor, after the refinement of the stability constants, exceeds R_{lim} the agreement shall be considered not satisfactory and other hypotheses about the chemical model should be tried.

For different hypotheses concerning the number and the type of the species formed in the complex chemical equilibrium, different values of R will be obtained. Using these values, the R factor ratio test [9] (otherwise called the Hamilton test) can be performed. Let us assume that the minimum value of the R factor, R_0, has been reached for the hypothesis H_0. An alternative hypothesis, H_i, which led to a value of R factor R_i, can be rejected at the α significance level, if

$$\frac{R_i}{R_0} > \mathcal{R}_{p,n-p,\alpha}$$

where p is the number of unknown parameters refined and $n-p$ is the number of degrees of freedom of the least squares adjustment. A table of significant values of $\mathcal{R}$ for different values of p, $n-p$ and α has been reported in ref. 9. If $n-p$ is large, this quantity may also be calculated according to

$$\mathcal{R}_{p,n-p,\alpha} = \sqrt{\frac{\chi^2_{p,\alpha}}{n-p}} + 1$$

Logarithms of the formation constants of the Cu^{II}-Me_7tetren complexes obtained using different hypotheses [a]

Hypothesis	$Cu(H_2Me_7tetren)^{4+}$	$Cu(HMe_7tetren)^{3+}$	$Cu(Me_7tetren)^{2+}$	$Cu(OH)(Me_7tetren)^+$	R
H_0	24 613 ± 0 005	20 070 ± 0 013	12 319 ± 0 035	3 037 ± 0 033	0 00068
H_1		20 137 ± 0 077	12 819 ± 0 892		0 01429
H_2		20 137 ± 0 013	12 371 ± 0 573	3 098 ± 0 562	0 01250
H_3	24 613 ± 0 043	20 070 ± 0 135	12 777 ± 0 465		0 00751

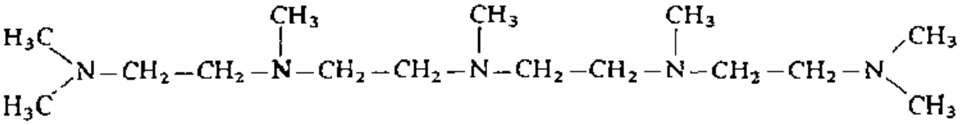
[a] Uncertainty intervals associated with all values are the standard deviations

C. RESULTS

The following examples will help to illustrate the practical application of the procedure described above.

(i) Copper(II) + Me$_7$tetren

Me$_7$tetren is a pentadentate ligand, of formula



$$(H_3C)_2N-CH_2-CH_2-N(CH_3)-CH_2-CH_2-N(CH_3)-CH_2-CH_2-N(CH_3)-CH_2-CH_2-N(CH_3)_2$$

which forms a number of complexes with the copper(II) ion. This system was the subject of a previous investigation[10] and the chemical model adopted on that occasion is shown in the schematic reaction diagram

$$Cu(H_2Me_7tetren)^{4+} \rightarrow Cu(HMe_7tetren)^{3+} \rightarrow Cu(Me_7tetren)^{2+} \rightarrow Cu(OH)(Me_7tetren)^{+}$$

The values of the logarithms of the formation constants obtained using this chemical model (hypothesis H_0) are reported in Table 1. The R factor is well below the value of R_{lim} and so we can be confident that the postulated model is a good representation of the experimental data. Other trials have been performed (hypotheses H_1, H_2 and H_3), assuming that one or two of the postulated complexes would not be present. The values of R which we have obtained are all higher than both R_0 and R_{lim}. Furthermore, in the most favorable case, the ratio R_3/R_0 (about 1.2) is greater than the Hamilton test value at the 0.05 significance level, $\mathcal{R}_{214,101,0.05} = 1.862$. At this level it is reasonable to reject all the alternative hypotheses and to conclude that the chemical model assumed is most probably the best solution of the complex formation equilibrium. These results can be shown more clearly in Fig. 1, where the corresponding R factor diagram has been reported. H_0 is considered to be the unique solution of our system.

(ii) Copper(II) + NNN′N′-tetramethylethylenediamine

The complex formation equilibrium in aqueous solution between the copper(II) ion and *NNN′N′*-tetramethylethylenediamine (Me$_4$en) has recently been investigated.[11] Several hypotheses about the chemical model for this system have been put forward, considering the possible formation of numerous complex species. The best fit of experimental data was obtained assuming that the following six complexes were present in the equilibrium mixture, viz. $Cu(Me_4en)^{2+}$, $Cu(OH)(Me_4en)^{+}$, $Cu_2(OH)_2(Me_4en)_2^{2+}$, $Cu_2(OH)_2(Me_4en)^{2+}$, $Cu_3(OH)_4(Me_4en)_2^{2+}$ and $Cu(OH)_2(Me_4en)$.

The value of the R factor calculated for such a hypothesis (R_0 = 0.00141) is well below the estimated value of R_{lim}, 0.00350. As shown in Fig. 2, five other trials on different

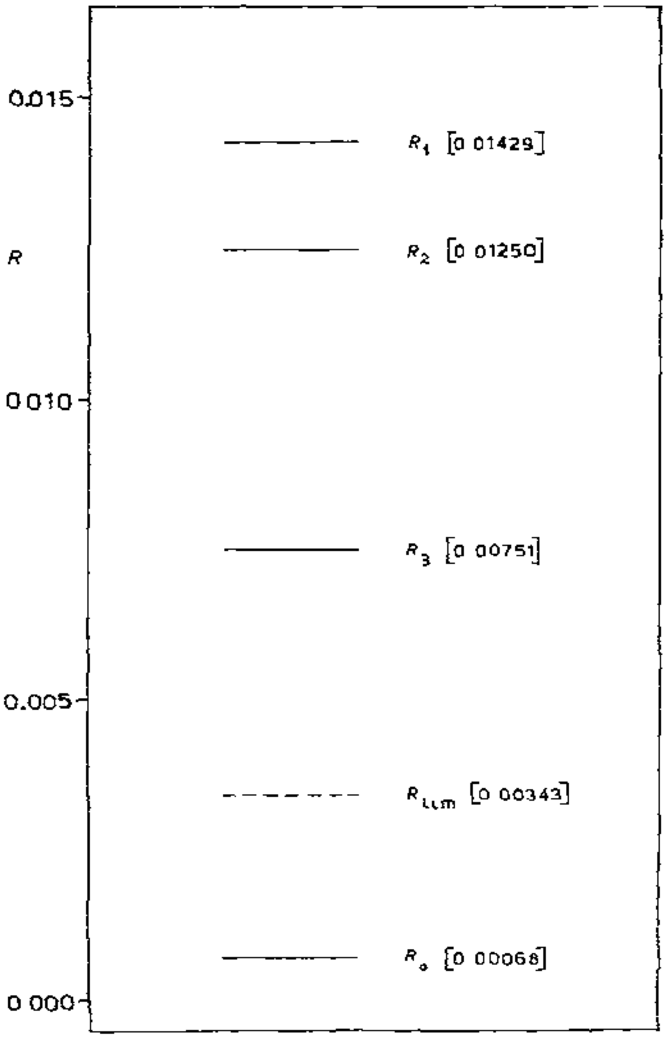


Fig. 1. R factor diagram for the system copper(II) – Me$_7$ tetren

chemical models led to R factors lower than $R_{\lim}$. Furthermore, the value of the Hamilton ratio for this system, $\mathcal{R}_{308,145,0.05}$, is equal to 1.848 and, using the Hamilton test, none of the five alternative hypotheses may be rejected. From the mathematical point of view, H_0 is the best solution; from the statistical point of view, there is no significant difference among all these hypotheses. As a consequence, the choice of the chemical model for this system is a matter of personal judgment.

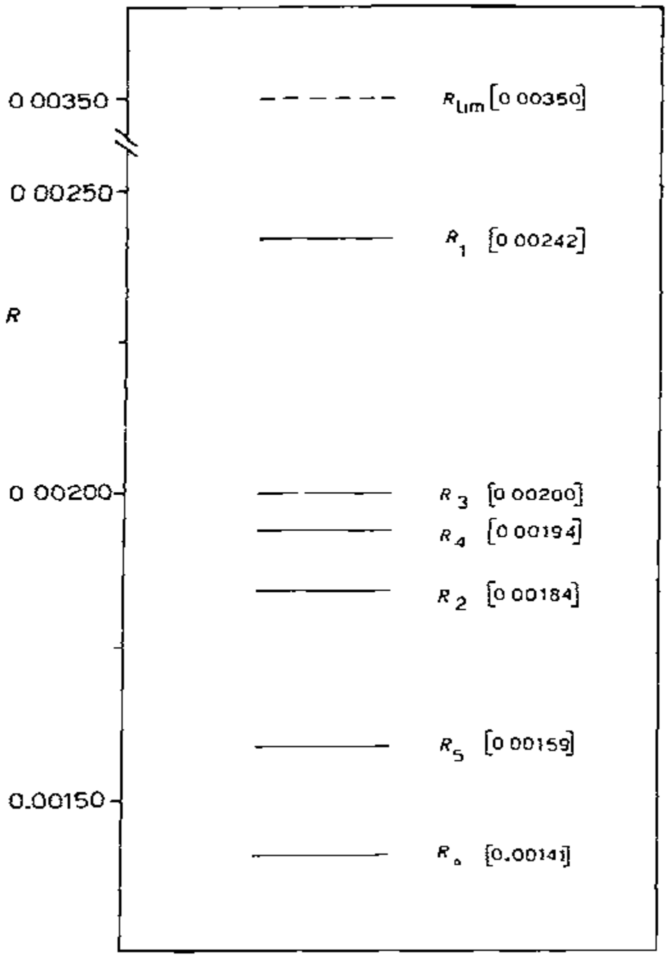


Fig 2 *R* factor diagram for the system copper(II) – Me_4en

D CONCLUSION

The values of the Hamilton test calculated for this type of system range between 1 8 and 2 0 These are rather high values, because, on this basis, hypotheses are to be rejected only if the corresponding R factors are at least twice R_0, and, as a consequence, non-unique solutions may often occur This is due to the fact that the number of degrees of freedom of the system is much lower than the number of parameters. The least squares adjustment is not so *overdetermined* as would sometimes be necessary. The availability of other types of additional experimental information, such as measured values of the concentration of the free metal ion and/or the free ligand, would surely be of help in these particular situations

REFERENCES

1 F J C Rossotti, H S Rossotti and R J Whewell, *J Inorg Nucl Chem* 33 (1971) 2051,
C W Childs, P S Hallman and D D Perrin, *Talanta*, 16 (1969) 1119
2 J Rydberg, *Acta Chem Scand* 15 (1961) 1723,
R S Tobias and M Yasuda, *Inorg Chem* 2 (1963) 1307,
D D Perrin and I G Sayce, *J Chem Soc A* (1967) 82,
I G Sayce, *Talanta* 15 (1968) 1397
3 L G Sillén, *Acta Chem Scand* 16 (1962) 159
4 L G Sillén and A E Martell Stability constants, *Chem Soc Spec Publ* , No 17, 1964,
A Vacca, D. Arenare and P Paoletti, *Inorg Chem* , 5 (1966) 1384.
5 J B Scarborough, *Numerical Mathematical Analysis* John Hopkins Baltimore, 1930, p 187
6 T G Strand, D A Kohl and R A Bonham, *J Chem Phys* , 39 (1963) 1307.
7 H L Steinmetz, *Ind Eng Chem* 58 (1966) 36
8 P Gans, *Chem Commun* , (1970) 1504
9 W C Hamilton *Statistics in Physical Sciences* Ronald Press New York, 1964
10 A Vacca *Ric Sci* 36 (1966) 1363
11 E Arenare P. Paoletti A Dei and A Vacca, *J Chem Soc* , in press